

To achieve full marks you are required to complete this assignment using R Markdown/Quarto to compile a **reproducible PDF** file for your submission and use the Git version control. On iLearn you only need to submit your pdf file, no need to submit your .Rmd/.Qmd file. For the Git repository you need to submit both files.

You need to submit your assignment via the provided submission link on iLearn by the due date. To further score marks for Question 4, you have to **push** the assignment file to provided Github repository.

You may discuss the assignment in the early stages with your fellow students. However, the assignment submitted should be your own work.

The R Markdown ‘Cheatsheet’ from the RStudio team is given [here](#).

In your answers to the questions below, produce the appropriate R output and explanation of the steps and results. Don’t include any more R output than necessary and include only concise explanations.

Rubric

This assignment includes two parts. The Part I practice assignment is worth 5%, and the part II main assignment is worth 20% of the unit marks. This is an assessment task that will test both your statistical knowledge and technical skills used in this unit.

Question 1 [25 marks] - Tests your applied Statistics skills

Question 2 [30 marks] - Tests your applied Statistics skills

Question 3 [10 marks] - Tests your RMarkdown technical skills

Marking Guide/Rubric for this question:

- Only 5 marks if the assignment file is compiled from RMD/QMD to HTML/Word (even if you then convert it into a PDF file).
- Full 10 marks if the assignment file is compiled directly from RMD/QMD to PDF with L^AT_EX.

Question 4 [5 marks] - Tests your Git version control technical skills

Marking Guide/Rubric for this question:

- Only 2 marks if only you uploaded to the designated repository once or didn’t include your StudentID as part of the assignment filename.
- Full 5 marks if you used a proper Github submission workflow: i.e., submitted at least twice into the designated repository with proper description, commit and push. Also, your StudentID is included in the assignment filename.

You should prioritise Questions 1 and 2 and present them in a format that you are comfortable with. Using Markdown/Quarto, \LaTeX , and Github can require substantial investment in your time and effort.

A small tutorial on R Markdown

The following are some notes to kickstart into your R Markdown journey (we discuss some of these in more details in Week 6 Part B Lecture and also in various SGTAs).

1. If you see an error message of `pdflatex not found`, then you are at the right place. To knit to a pdf you need to install \LaTeX on your computer. This is rather big (e.g. MacTeX is approximate 4.7Gb and MiKTeX 192Mb), but a recommended option. Before installing anything, make sure you have admin right to your computer before you start. If you have encountered issues with the installation of \LaTeX , then you could try to install via `tinytex` which is much more light-weight. Open R and enter the following commands:

```
install.packages("tinytex")
tinytex::install_tinytex()
```

- See Week 6 Part B Lecture for some other alternatives.
 - For Mac users, you may be asked to install Xcode (another rather big installation). We only need a small piece of it called the `command-line tools`. Run the following line: `xcode-select --install` in the `Terminal` to continue. You should be able to find the `Terminal` tab next to the `Console` tab in RStudio.
2. To communicate your assignment results to us (this is the knitting part), you need to know some markdown & \LaTeX syntax. Learning Markdown syntax will help with your formatting while learning \LaTeX syntax will allow you to typeset Mathematics (copying β into your `.Rmd` file and assuming it would work is one of the most common errors) in your assignment. Here are some resources to get you started.
 - Markdown tutorial - 10 minutes tutorial [link](#)
 - Mathematics in R Markdown [link](#)
 - Remember Google is your best friend, and you should google whatever error messages you got. Able to debug your own code with Google (learn how to select the right keywords to improve your searches) and by trial and error is part of the learning process. Please give this a go before reaching out for help.
 - Now create a new R Markdown document from RStudio and knit it.
 3. If you are experiencing persisting or last minute (\LaTeX) compiling issues, Posit Cloud is an excellent platform. Simply upload everything online and knit.
 4. For those who decide to use the Posit Cloud platform, you will have to **download** the pdf file instead of printing to a pdf at the end. Printing to a pdf, unfortunately, will turn each page into an image and then the submission system will reject it.
 5. It is also our recommendation to knit often so that you know which line(s) of code is(are) giving you the problem. (There is a keyboard short-cut for knitting.) This is not so dissimilar to when you work with the console and only run a line at a time to identify the issue.
 6. Another common mistake is that students use the code `read.csv("dat.csv")` and then assume R would be able to know (magically) that you are referring to `dat.csv` in a folder far far away (in the `Download` folder probably) from your `.Rmd` file. At this point of the semester, you should all have your `.Rproj` file and workspace setup already so that everything will be run from there. Please go back to Week 1 lecture for more details.

7. If you are stuck, create a post on the iLearn forum! Also, check earlier posts before creating a new one. Most of the time, your issues have been discussed and resolved already.

Instructions for Git version control

To score marks in Question 4 you need to pull the assignment file from the repository, make changes to the template RMD file, compile it to a PDF file, stage the changes, add proper description (Summary and Description) and push the file to the repository. Do it at least twice to demonstrate your level of skill in version control workflow. Refer to the following link to find out how Git version control works in RStudio.

- Happy Git and GitHub for the useR [link](#)
- RStudio Support blog article [link](#)
- Posit Cloud users [link](#)

Once RMarkdown and Git are installed (and RStudio is configured for both) on your laptop, open the following Github repository link provided on iLearn.

1. Accept the invitation and wait until you received a confirmation email.
2. In RStudio open New project, and choose Version Control, then choose Git.
3. Copy the repository URL, eg.
<https://github.com/MQ-STAT2170-STAT6180/2024-s1-stat2170-stat6180-assignment-yournamehere>
[this link won't work as it is just an example]
 - you may add an exact folder location on your laptop,
 - when you create the project, the files will be downloaded automatically, i.e. the pull request will clone the repository on your laptop.
4. **Assignment-your_name_studentID_here.Rmd** is your starter file for your answers in Rmarkdown. We strongly recommend you to rename it with your details.
5. Open the (renamed) starter file.
6. In the default RStudio layout, you should be able to find the **Git** section in your top right-hand side window (alongside **Environment** and **History**) to **stage** updated/changed files - please remember to add proper **description**.
7. When you click on a **Push** button, the **staged** files (RMD and PDF) will be uploaded to your repository. If this is your first **Push**, you should log onto **Github** to check your files have been uploaded properly.

Question 1 [25 marks]

Economists are exploring the relationship between **Consumer Confidence Index** and the change in retail sales in 243 different cities. The dataset `sales.csv` contains information about the following variables.

Variable	Description
Index	Consumer Confidence Index
Sales	Change in retail sales (in percentage)

- [3 marks] Load the data from the file `sales.csv` and create a scatter plot of **Sales** against **Index**. Comment on the relationship between the two variables.
- [5 marks] Fit a simple linear regression model and named it as **M1** to predict **Sales** using **Index**. Validate the model through diagnostic checks and comment.
- [4 marks] Fit two polynomial models of order 2 and order 3 to predict **Sales** using **Index**. Name the quadratic model as **M2** and the cubic model as **M3**. There is no need to validate the polynomial models at this stage. Compare and comment.
- [5 marks] Plot the data and add the three predicted lines from models **M1**, **M2**, and **M3** to your plot. Comment on the fit of the models.
- [4 marks] Assess the significance of the linear, quadratic and cubic terms in **M3** using a Sequential Sum of Squares (Hint: A sequential ANOVA is necessary.) Comment on the results.
- [4 marks] Choose the best model among **M1**, **M2**, and **M3** and validate it. Provide reasoning and comment on the model fit.

Question 2 [30 marks]

In the rapidly evolving landscape of marketing, companies are constantly seeking strategies to capture consumer attention and foster brand loyalty. The effectiveness of these marketing campaigns can be influenced by a myriad of factors, including the medium of the campaign and the geographical region in which it is deployed. Your task is to determine the impact of different marketing campaign types on customer engagement across various regions. The dataset `campaign.csv` contains information about the following variables.

Score	The percentage increase in engagement score per customer
Region	The geographical region where the marketing campaign was conducted: Urban or Rural
Type	The type of marketing campaign used: Social Media, Email, or Billboard

- a. [5 marks] Construct two different preliminary graphs that investigate different features of the data and comment.
- b. [6 marks] Write down the **full** interaction model for this situation, defining all appropriate parameters.
- c. [8 marks] Analyse the data to study the effect of **Type** and **Region** on the percentage increase in engagement **Score** at 5% significance level. Remember to
 - state the null and alternative hypothesis for each test,
 - check assumptions (model diagnostics),
 - interpret the results.
- d. [7 marks] Repeat the above test analysis for the main effects.
- e. [4 marks] Using TukeyHSD produce multiple comparisons between each level for both **Type** and **Region**. Comment on the effectiveness of the marketing campaign **type** on customer engagement **scores** and also the impact of **region** on customer engagement **scores**. (Hint: Confirm the design is balanced before proceeding with the TukeyHSD test.)